DOCUMENT RESUME

ED 279 186 FL 016 494

AUTHOR Altenberg, Bengt
TITLE Lexical and Sex-Related Differences in Spoken and Written English: Some Results of Undergraduate Research at Lund University.
PUB DATE 84
NOTE 21p.; In: Ringbom, Hakan, Ed,; Rissanen, Matti, Ed. Proceedings from the Nordic Conference for English Studies (2nd, Hanasaari/Hanaholmen, Finland, May 19-21, 1983); see FL 016 498.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Comparative Analysis; *English; Foreign Countries; Higher Education; Language Styles; *Language Usage; Language Variation; *Oral Language; Phrase Structure; Research Papers (Students); *Research Projects; Research Skills; *Sex Differences; Vocabulary; Word Frequency; Writing Instruction; *Written Language
IDENTIFIERS *Lund University (Sweden)

ABSTRACT

A number of studies written by Swedish students of English in a university research and writing skills program focused on lexical and sex-related differences between spoken and written English. The lexical studies all confirmed the existence of two divergent modes of communication, a spoken "verbal" style and a written "nominal" style. The reports show the contrasts between the styles as systematic and fundamental, involving categories and structures at all levels of linguistic organization and closely related to the contrast between face-to-face conversation and public expository prose. The studies have confirmed and enriched each other from widely different viewpoints and agreed with the findings of other comparable studies. The studies of sex-related language differences have sometimes agreed with each other but have often given a confusing picture of non-significant or conflicting tendencies. They have shown that more research is necessary to disentangle the many sex-related variables in conversational behavior. However, the studies show that projects of this kind can be of great value and can contribute to the understanding of various aspects of spoken and written English. (MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED279186



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Enkvist

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Bengt Altenberg
Department of English
Lund University

# LEXICAL AND SEX-RELATED DIFFERENCES IN SPOKEN AND WRITTEN ENGLISH
## SOME RESULTS OF UNDERGRADUATE RESEARCH AT LUND UNIVERSITY

### 1. BACKGROUND

When the project English in Speech and Writing (in Swedish Engelska i tal och skrift - ETOS for short) was started at the English Department, Lund University, in 1981, one of its aims was to incorporate undergraduate students into the research activity, for the mutual benefit of the students and the project itself (see Tottie et al 1983). Since then, more than twenty-five undergraduate studies - mainly by third-term students - have been produced on various aspects of spoken and written English (see Altenberg & Tottie, this volume). The aim and scope of these studies have of course been restricted, and the quality has varied, but many of them have given valuable contributions to our knowledge of English speech and writing. The purpose of this paper is to summarize some of the results that have emerged in two different areas: (a) vocabulary and (b) language and sex. My presentation, although brief and sketchy, will, I hope, illustrate the kind of undergraduate research that is in progress at Lund, and perhaps serve to inspire others to do similar work.

I have chosen to concentrate on lexical and sex-related differences, not because they are necessarily more central or fully covered than other areas investigated in the project, but because they have been the subject of partly overlapping studies which offer interesting opportunities for comparison, both mutually and with other studies, on English as well as Swedish material. They are also good examples of two radically different research areas that can be fruitfully investigated, even by linguistically inexperienced students, on a limited material.

Although they may appear to concern fairly superficial properties of language, both areas in fact reveal central features of spoken and written discourse and the way these are related to various situational constraints. The lexical studies focus on certain structural and systematic (rather

FL016494



BEST COPY AVAILABLE

than individual) differences in the vocabulary of speech and writing, such as the distribution of word-classes and patterns of word-formation. Apart from their obvious stylistic interest (eg nominal vs verbal style, simple vs complex words), differences in these respects are also grammatically revealing, since they reflect syntactic options (eg nominalization vs finite expression, compression vs dilution) which are closely associated with the different planning and processing conditions of spoken and written discourse.

The studies in language and sex have a strong sociolinguistic bias and are primarily concerned with interactive and paralinguistic features in spontaneous conversation. Although their main purpose is to establish sex-related differences, they also highlight phenomena that play an important role in conversation regardless of sex (turn-taking, hesitation, tag questions, softeners, etc) and which are determined by a range of rather elusive factors.

The material used for these studies (described in detail in Tottie et al 1983) consists of surreptitiously recorded spontaneous conversations from the London-Lund Corpus of Spoken English (printed in Svartvik & Quirk 1980) and public expository prose from the Lancaster-Oslo/Bergen Corpus of British English (described in Johansson et al 1978). These can be regarded as 'canonical', maximally contrasted varieties of spoken and written English, which are likely to reveal differences of great linguistic and pedagogical interest. For student use, two smaller subsets were selected from these corpora: a 'Midi' corpus of 100 000 words, and a 'Mini' corpus of 10 000 words, from each variety. In most cases the students have used the Mini corpus, but sometimes larger samples have been necessary. It goes without saying that the material is often too small for far-reaching conclusions, but for comparative purposes even small-scale investigations based on this material are of considerable interest.

## 2. SPOKEN AND WRITTEN VOCABULARY

Three studies were devoted to the distribution of lexical categories and other general properties of the vocabulary in the two corpora. The following aspects were dealt with:

- o word-class distribution
- o word-formation (especially derivation)
- o lexical density (the proportion of content words)

- o lexical variation (type-token ratio)
- o lexical originality (text-specific vocabulary)
- o rank-order differences of high-frequency words

Some of the findings, obtained by random sampling from the Mini corpus, are outlined in Table 1. The table gives the relative frequency of the features examined, expressed in tokens per 100 words. A rough picture of the tendencies in the material is given in the right-hand column, where a '+' indicates predominance in speech, a '-' predominance in writing, and a '=' equal, or almost equal, distribution. No significance testing has been attempted, and small differences between the two varieties should therefore be treated with caution.

## 2.1. WORD-CLASS DISTRIBUTION

The distribution of word-classes in the two corpora was examined in a study by Stina Ohlsson. Her results, confirmed and amplified in several respects by the other studies to be discussed below, reveal a clear differential pattern: the spoken material has a higher frequency of lexical and auxiliary verbs, pronouns and interjections, and the written material a higher frequency of adjectives, determiners, nouns and prepositions. This alignment is in accord with Wells's (1960) well-known distinction between a 'verbal' and a 'nominal' style, based on the (impressionistic and consciously simplified) observation that language use is characterized by two structural alternatives that can be paired and transmuted. 'Verbal' forms of expression generally have a 'nominal' counterpart (and vice versa), and the change from one to the other tends to affect related word-classes in a systematic way: verbs attract adverbs and conjunctions, nouns attract adjectives and prepositions (cf because his father suddenly died and due to his father's sudden death).

Wells's distinction, which was largely based on writing and not explained in situational terms, can be profitably applied to spoken and written discourse, and be clarified with reference to the different situations in which these are produced (on situational differences, see eg Ochs 1979, Akinnaso 1982, Chafe 1982). Speech, especially in its prototypical form - spontaneous face-to-face conversation - is likely to promote a 'verbal' style for a number of reasons. Its speed of delivery and real-time planning offer little opportunity for editing and integration of ideas into large and complex structures, but break down the flow of

Table 1. Some lexical differences in the spoken (S) and written (W) material.

| Categories | S | W | Diff |
|---|---|---|---|
| Word-classes: | | | |
| Nouns | 12.5 | 28.5 | - |
| Verbs | 27.4 | 12.4 | + |
| auxiliaries | 7.2 | 5.2 | + |
| lexical verbs | 20.2 | 7.3 | + |
| finite | 13.5 | 5.2 | + |
| non-finite | 6.7 | 1.6 | + |
| Pronouns | 21.2 | 8.3 | + |
| Prepositions | 9.1 | 13.0 | - |
| Adverbs | 11.1 | 9.3 | + |
| Conjunctions | 9.1 | 9.8 | = |
| Determiners | 1.9 | 11.9 | - |
| Adjectives | 3.8 | 6.7 | - |
| Interjections | 3.8 | 0.0 | + |
| Word-formation: | | | |
| Simple words | 93.5 | 86.5 | + |
| Compounds | 0.5 | 3.5 | - |
| Derived words | 6.0 | 10.0 | - |
| Prefixation | 0.6 | 0.9 | = |
| Suffixation | 5.4 | 9.1 | - |
| Lexical density | 39 | 46 | - |
| Lexical variation | 62 | 70 | - |
| function words | 50 | 49 | = |
| content words | 81 | 95 | - |
| Lexical originality | 77 | 80 | - |

information into a succession of short, grammatically simple, finite units; it takes place 'in context', which means that much information is shared by the participants and can thus be taken for granted, or else be clarified by deictic or paralinguistic means; it is both egocentric and

sociocentric, which promotes expressions related to the contact, roles, feelings and beliefs of the participants; finally, it is naturally message-oriented with little regard for verbal form and correctness.

Writing, on the other hand, especially the public, expository kind selected for the ETOS studies, is produced with time for planning and editing, which permits an organized way of expression and the integration of ideas into larger and more complex structures; it takes place 'out of context', generally without a clearly identifiable audience, which means that assumptions, references and ideas have to be made explicit and 'contextualized' by linguistic (but non-prosodic) means; it is a public activity, resulting in a permanent record and shaped by a long tradition, which makes for 'channel-orientation' and a high degree of style-consciousness. All these factors promote a 'nominal' style characterized not only by a greater number of (often abstract) nouns and nominalized verbs and adjectives, but also by other word-classes associated with the noun phrase, such as determiners, attributive adjectives and prepositions.

It is also instructive to look at the degree of difference shown by the distribution of the word-classes. Interjections only occur in the spoken sample; they are typically speech-specific and need no further comment. The second largest difference is to be found among the determiners, which are about six times as common in the written as in the spoken sample. This is largely due to an increase in the use of the definite article in the written material, and can be explained as the combined result of the nominalizing tendency in written discourse and the fact that reference is often achieved by pronouns in speech.

On the whole, the different patterning of nouns, pronouns and verbs in the two corpora shows an interesting correlation. The increase of nouns in the written material is counterbalanced by a corresponding increase of verbs in the spoken material, a contrast which is well in line with Wells's argumentation for nominal/verbal commutability. The fact that the verbal increase in speech is primarily due to a greater employment of lexical (rather than auxiliary) verbs further supports this. However, the higher frequency of nouns in the written sample is also connected with the increase in pronouns in the spoken sample. As mentioned, the 'decontextualized' writing situation requires greater explicitness and referential clarity. This is to a large extent accomplished by endophoric referring expressions consisting of a definite determiner and a noun phrase. In face-to-face conversation, on the other hand, reference is either evident from

the speech situation, or else achieved by paralingustic means or by deictic pronouns (cf Ochs 1979:68). Moreover, the disintegrated character of spoken discourse makes heavier demands on the use of anaphoric pronouns for cohesive purposes, to link referents in short successive information units. As Thavenius has shown (this volume), this results in a higher frequency of pronominal 'chains' in spoken texts. Another speech-specific phenomenon is the profusion of mental self-reference expressions and interactive 'softening' particles (I suppose, I think, I mean, you know, you see, etc), which serve a variety of pragmatic and stylistic functions in face-to-face conversation. All these factors contribute to a higher frequency of pronouns (and finite verbs) in spoken discourse, especially the speech-role pronouns you and I and the third-person pronoun it (cf Thavenius, this volume). This is also confirmed by Nordin's study (see below), in which 76% of the pronouns in the spoken sample were personal, as against 29% in the written sample. Hence, the differential patterning of nouns, pronouns and verbs in the material is hardly surprising (for similar results, see DeVito 1966, Poole & Field 1976, Chafe 1982 and, for Swedish, Einarsson 1978:42ff); it underlines the close correlation between word-class distribution and situational/pragmatic factors.

The differential patterning of the remaining word-classes is less pronounced. The moderate increase in adjectives in the written material is partly offset by a greater frequency of adverbs in conversation, but an additional reason is that adjectives are about twice as common in attributive as in predicative position in the written sample, while the distribution is even in the spoken sample. This confirms several earlier observations that the use of attributive adjectives is above all a written phenomenon (cf O'Donnell 1974, Kroch & Hindle 1982:36 and, for Swedish, Einarsson 1978: 123ff), a circumstance which is probably related to the greater opportunity for planning and elaboration permitted by this medium.

The conjunctions are the only word-class that does not reveal a significant distributional difference in the two corpora. This is somewhat surprising, as it conflicts with the otherwise very neat 'Wellsian' patterning of the results, as well as with certain previous studies. One reason for this deviation is probably the different functions of coordinating and subordinating conjunctions. No such distinction was made in Ohlsson's study, but she comments on the great frequency of and in the spoken sample. The preference for 'loose' coordinated structures in spoken discourse is well attested (cf Leech & Svartvik 1975:159, Chafe 1982) and it is likely that a

separation of the two types of conjunction would have been clarifying. This does not necessarily mean that there is a neat symmetric opposition between coordination in speech and subordination in writing. Although subordination is generally said to be more characteristic of written discourse, there is evidence that the situation is more complex (see eg Poole & Field 1976). Many subordinators like that, if and because are very frequent in spoken discourse (on because, see Altenberg, forthcoming, and on Swedish equivalents, Einarsson 1978:108), and it is obvious that more research is needed to clarify the use of subordination in the two varieties.

A few words should also be said about the verbal subclasses. Two broad distinctions were made in Ohlsson's study: between lexical and auxiliary verbs, and (among the lexical verbs) between finite and non-finite verbs. All these subclasses had a higher frequency of occurrence in the spoken material, which underlines the 'verbal' character of conversation. The only category that might have been expected to deviate from this pattern is the non-finite subclass, which is often said to be typical of written discourse (eg DeVito 1967). However, the discrepancy here, as in the case of conjunctions, is probably due to differences in descriptive delicacy. It is evident that a distinction must be made between infinitive constructions, which are common in both varieties, and different participle constructions, some of which are more typical of writing (cf O'Donnell 1974, Leech & Svartvik 1975:161, Chafe 1982). Studies intended to explore syntactic differences of this kind in the ETOS material are in progress.

## 2.2 WORD-FORMATION

From what has been said above it is obvious that the distribution of word-classes in the two corpora tells us more about their syntactic (and situational) properties than about their vocabulary. To a large extent this is also true of word-formation. As has often been pointed out, word-formation is at the crossroads of morphology and syntax, and spoken-written differences in this area are therefore likely to have considerable grammatical relevance. This is demonstrated in a study by Eva Ryberg, based on random sampling from the Mini corpus. Ryberg makes an initial distinction between two major types of word-formation - compounding and affixal derivation (conversion and other minor possibilities are disregarded) - but her main concern is with the latter type.

As shown in Table 1, the incidence of compounds and derivatives is very low in the material (partly due to Ryberg's strict definition of deriva-

tion), but the proportions are higher in the written (4% and 10%, respectively) than in the spoken sample (1% and 6%). Hence, although the differences are not overwhelming, there is a stronger preference for simple words in speech and derived words in writing, a polarization which can be related to the traditional division of English vocabulary into two lexical strata, variously described in oppositional terms like short-long, native-foreign, common-rare, informal-formal (cf Vachek 1980:273ff, Akinnaso 1982: 100f).

Of the two means of derivation - prefixation and suffixation - the latter dominates greatly in both varieties, both in terms of types and tokens (91% vs 9%). This tendency does not differ significantly in the two varieties. A more notable difference (not shown in the table) is that, while suffixes are more frequently used to form nouns in writing (again in terms of types as well as tokens), the adverb-forming suffix -ly predominates in speech. Other types of suffixes (forming adjectives and verbs) are rarer and more evenly distributed in the two corpora. The same applies to the use of prefixes (types and tokens), which show few distributional differences between the two varieties. In both, the negative forms un- and in- are most frequent.

Although these results do not reveal a consistent differential pattern, the divergent uses of the noun- and adverb-forming suffixes confirm the grammatical tendencies already demonstrated by the word-class distribution in the two corpora. The greater frequency of ly-adverbs in conversation correlates not only with the 'verbal' character of impromptu speech, but also with the greater personal and emotional involvement of speakers. This is borne out by the profusion of intensifying expressions (eg it was absolutely illogical, it's awfully tempting) and by the four most frequent ly-adverbs in the spoken sample: actually, really, entirely and terribly.

Conversely, the predominance of noun-forming suffixes in the written sample underlines the 'nominal' character of this variety. The frequency of nouns ending in -tion, -ment, -ness, -ity, etc, highlights the written tendency towards abstractness and detachment, or what Akinnaso (1982:112) calls "the abstraction of a statement from temporal, spatial, and situational boundaries". Obviously, other situational factors contribute to differences of this kind, for example topic and degree of formality. As Ryberg observes, when the participants in one of her spoken texts "turn from tea-making to discussing lectures, the proportion of derived nouns increases considerably". Yet, there is no doubt that the different oppor-

tunities for planning and 'contextualization' are an important channel-dependent reason for the differences noted above. Even a fairly formal oral discussion of a learned topic would hardly be couched in the 'nominal' terms characteristic of written examples like the following:

> Coleridge's definition of the secondary imagination, with its stress on the transmutation of experience by an essentially creative process into something of visionary freshness, can be taken as an exact description of Lawrence's art; (LOB: C 13.21-23)

## 2.3 LEXICAL DENSITY, VARIATION, ORIGINALITY AND RANK ORDER

Tendencies of a similar kind are also reflected in the four lexical properties examined in a study by Henrik Nordin: lexical density, variation and originality, and certain rank order differences in the use of high-frequency words.

By lexical density (LD) is meant the percentage of content (open-class) words in a text. This measure has been used by Ure (1971) to account for her observation that 'language-in-action' (ie language that accompanies an activity and relies heavily on features in the speech situation to be intelligible) contained a higher percentage of function (closed-system) words with a purely grammatical or pragmatic function than other varieties. Since, by definition, content words convey lexical meaning and serve to carry the information forward in a way that function words do not, a high LD can be said to indicate a high density of information. Not surprisingly, the strongest contrast in Ure's material was found to exist between interactive speech (with an LD never exceeding 36%) and written informative texts (with an LD always exceeding 44%). Differences of a similar kind have also been shown to distinguish the written production of native speakers and Swedish learners of English (see Linnarud 1979:222).

Nordin's application of Ure's measure to the ETOS material confirms her findings, with the exception that the spoken sample has a somewhat higher average LD (39%) than Ure's study would predict (36%). Nevertheless, there is a notable difference in LD between the written (46%) and spoken samples (39%). This supports the observation made above about the structural organization of the two varieties: speakers tend to space out information and rely on various contextual cues in signalling their intentions, using function words (especially pronouns) to a greater extent than writers. Conversely, the latter have time to plan and pack more information into their discourse. Also, they have to rely on the linguistic channel in conveying

their intentions, and are therefore forced to greater explicitness and elaboration.

The second property, lexical variation (LV), expresses the type-token ratio of the words in a text, ie the ratio of the number of different words (types) to the total number of words (tokens). This is an old measure going back to psycholinguistic studies in the 1940s, and has been used in several quantitative investigations to show that written language is normally characterized by a more varied vocabulary than spoken language (see Akinnaso 1982:100). As shown in Table 1, the LV in the ETOS material confirms these earlier findings: the written sample has a higher LV (70%) than the spoken one (62%). However, the difference applies to the content words only, since no significant contrast was found in the use of function words. This is as might be expected: content words are open-class items permitting considerable variation, while the function words form a closed set of grammatically indispensable elements which can be varied very little. Hence, these results are not surprising. They emphasize the fundamental situational difference between the two varieties - the different opportunities for planning and the channel-oriented, formal character of public writing, contrasting with the message-oriented, repetitive and informal character of spontaneous conversation.

Lexical variation is in some ways related to the next property examined in the material, lexical originality (LO). The LO of a text can be captured by counting the proportion of text-specific words, ie words that are unique to a given text and not found in other compared texts. Although text-specific words are highly topic-dependent, the proportion of such words are of interest because it can tell us something about the generality of a given vocabulary in relation to some selected norm or contrast (eg common (easy) vs rare (difficult) words). The measure does not seem to have been used much in studies of spoken and written discourse, but can be related to the greater lexical predictability claimed for spoken discourse and to Bernstein's distinction between restricted and elaborated codes (cf Akinnaso 1982:115).

Nordin's study shows that, on the average, the written texts in his ETOS sample contain more specific words than the spoken ones (80% vs 77%). The difference is not very great, however, and there is a great deal of variation between the sample texts. For obvious reasons the text-specific words were largely confined to content words, and the shared words to func-

tion words (the only shared content words in the two samples being common relational verb forms like have, is, was, get).

To supplement these rather general results Nordin made a closer study of the hundred most frequent words in the entire LLC and LOB corpora, based on the rank-lists presented in Svartvik et al (1982:43-49). For natural reasons, the majority of the word forms (68 items) appearing in these lists are identical. Most of these 'shared' items are function words, but there are also a few all-purpose content words (said, people, time). However, 32 word forms occur only in one of the rank-lists. All of these are confined to the LLC list, which suggests that conversational English is marked stylistically by a fairly small set of high-frequency items. Apart from obviously speech-specific forms like interjections (oh), minimal responses (yes, yeah), supports ([m̆], [m̄]), filled pauses [əm], [ə:m]) and contractions (it's, don't, that's, I'm, etc, which are counted as single words in the lists, and further described in Forsheden 1983), these include a score of 'ordinary' words of great interest. The most frequent are (in rank order): know, think, mean, got, see, sort, just, because, really, get and thing. Most of these reflect conversational features already mentioned in the preceding sections: the frequent use of mental self-reference expressions and 'softeners' (you know, you see, I mean, I think), emphatic particles (just, really), terms of vagueness (sort of, thing) and explanatory conjunctions (because).

A third category, finally, consists of items that occur in both lists, but with a considerably higher rank in the spoken one. The ten most dramatically 'rank-shifted', speech-favoured items are: well, very, do, people, you, what, so, no, like and I. Although some of these have several functions, their conversational relevance is easy to see: the speech-role pronouns I and you, the question-response items do, what and no, the intensifiers very and so (the latter also a speech-promoted result conjunct; see Altenberg, forthcoming), the initiater-particle well (see Svartvik 1980) and the multi-purpose words people and like.

The three lexical studies summarized above have been based on a rather restricted material and have focused on seemingly rather distinct properties of English vocabulary. Yet, together they give a fairly clear and, I think, reliable picture of certain fundamental differences between spoken and written English. The various aspects examined - the distribution of word-classes, word derivation, lexical density, variation and rank order - all point to the existence of two divergent modes of communication, a

spoken 'verbal' style and a written 'nominal' style. What distinguishes these styles cannot be described in purely lexical terms, however. The distinction involves options of many kinds, at all levels of linguistic organization, from the choice of particular words and syntactic constructions to the 'chunking' and linking of information in coherent discourse. Hence, quantitative lexical studies, even of the restricted kind reported here, can be very rewarding and serve as a starting-point for explorations of differences at other levels. Moreover, they demonstrate with great clarity the close connection between spoken and written communication and the situations in which these take place.

## 3. LANGUAGE AND SEX

Men's and women's language has attracted great interest in recent years (see eg Elert 1979, Kramarae 1982). Observed differences can generally be related to the social roles associated with the two sexes and to the linguistic attitudes and expectations that these tend to foster. As Trudgill (1974:101) puts it, men and women "speak as they do because they feel a particular kind of language to be appropriate to their sex". The spoken ETOS material is in many respects well suited for investigations of such differences, even if the influence of sex is not always easily isolated from other conditioning factors, such as age, occupation, participant relationship, idiosyncratic usage, etc.

Two partly overlapping ETOS studies of sex-differentiated language, conducted independently by two third-term students, Eva Klang and Carita Rosmark, will be reported here. Although restricted in aim and scope, these studies have a certain comparative interest. Klang based her investigation on same-sex conversations, whereas Rosmark used cross-sex conversations (the most widely studied variety). Since linguistic sex differences are largely role-determined, they can be expected to vary in character depending on the presence or absence of the opposite sex.

Some other facts about the material should be noted. Klang's same-sex study contrasts one male conversation between two middle-aged university teachers (LLC: S.1.4) with one female conversation between four academic secretaries aged 21-35 years, each text containing c 5 000 words. Rosmark's cross-sex study is based on three conversations (LLC: S.2.10, S.2.11b, S.2.13) totalling c 14 000 words. Each conversation involves a married couple (discounting one prompter of either sex) of varying ages (20-30 years) and occupations (two social workers, two BBC employees, one merchant

banker, one housewife). Apart from these variations in age and occupation, the conversations of both studies have several essential features in common: the language is surreptitiously recorded impromptu speech, the participants generally know each other well, and their background as well as the context of situation (and frequently the topic) can be broadly described as 'academic' in character.

In all, ten conversational features were investigated. These are presented in Table 2 under four headings: turn-taking, fluency, interaction signals and prosody. Features examined in one or both of the studies are distinguished by the students' initials: EK (Klang) and CR (Rosmark). For the sake of clarity only the main tendencies are shown. A feature exhibiting significant predominance with one of the sexes is indicated by a plus sign; non-significant tendencies are marked by parentheses and contradictory results by a question mark.

As shown in Table 2, the men in the material tended to be more talkative: they produced a significantly larger amount of speech in the cross-sex conversations, and they generally took longer turns (on the average 21 words as against 14 words for the women). The women, on the other hand, produced more but shorter turns. In both studies, the longest turns were normally produced by the men and the shortest by the women.

These results contradict the popular view of women as the garrulous sex, as reflected in many proverbs ("Where there are women and geese there's noise" - cf Kramarae 1982:87). They also accord well with the findings of recent sex-related studies, which have either revealed greater male talkativeness (Thorne & Henley 1975:16, Einarsson 1976:64) or non-significant differences (Oreström 1983:96ff). The results also suggest that the women have a tendency to be more interactive (in the same-sex conversations) or to adopt a supportive or listening role (in the mixed conversations). Conversely, although the men are apt to be verbally dominant (in quantitative terms), this does not make them significantly more 'interruptive' than the women, at least not in the same-sex conversations. Hence, the claim sometimes made that men interrupt more than women may either be restricted to cross-sex conversations (cf Zimmerman & West 1975: 115ff) or be greatly exaggerated (cf Oreström 1983:146ff).

Women have frequently been said to be more fluent speakers than men, with fewer signs of hesitation such as pauses, repetitions and false starts (see Einarsson 1976, 1978:148ff). This is only partly supported in

Table 2. Some sex-related features in spontaneous English conversation.

| FEATURES | MEN | ? | WOMEN |
|---|---|---|---|
| TURN-TAKING | | | |
| 1 Amount of speech (CR) | + | | |
| 2 Number of turns (EK, CR) | | | + |
| 3 Average length of turns (EK, CR) | + | | |
| 4 Interruptions (EK) | (+) | | |
| FLUENCY | | | |
| 5 Pauses (EK, CR) | | | |
| a) silent | | | + |
| b) voiced | + | | |
| 6 Stuttering (EK, CR) | | | |
| a) repetitions | + | | |
| b) false starts | | ? | |
| c) tongue slips | + | | |
| INTERACTION SIGNALS | | | |
| 7 Tag questions (EK, CR) | | ? | |
| 8 Softeners (CR) | (+) | | |
| 9 Laughs (CR) | | | + |
| PROSODY | | | |
| 10 Nuclear tone (EK) | | | |
| a) fall | | | (+) |
| b) rise | (+) | | |
| c) fall-rise | (+) | | |
| d) fall+rise | (+) | | |

the ETOS material. While the women produced a greater number of silent pauses, the men's speech contained more voiced pauses ([ə(m)], [ə:(m)], etc), repetitions and tongue slips. The distribution of false starts (lexical or syntactic reorganization) showed conflicting tendencies in the two studies.

These results reveal no consistent differential pattern, and it is not obvious how they should be interpreted. Hesitation phenomena have a number of reasons - emotional, cognitive, communicative and idiosyncratic. Gener-

ally, they are signs of mental processing (planning and replanning) and have been shown to correlate positively with factors like education and grammatical complexity (see Einarsson 1978:148ff). However, they can also be used as floor-holding devices. Although the silent pauses were by far the most common hesitation type with both sexes in the material, the male predominance of audibly marked hesitation features, especially voiced pauses and repetitions, might possibly be taken as evidence that the men tried to produce more complex language and/or to control turn-taking to a greater extent than the women. These possibilities were not examined in the material, however,

Women are often said to be 'submissive', 'hesitant' and 'uncertain' in mixed conversations (see eg Kramarae 1982:92ff). Three types of 'interaction signals' were selected to test this claim: tag questions, softeners (well, you know, you see, sort of, kind of, I mean) and laughs. According to R. Lakoff (1973:54), tag questions are characteristically used by women to avoid full commitment to a statement ("This is silly, isn't it?"), but the contradictory results from the ETOS material do not support this. The fact that tag questions predominated among the men in the same-sex material and among the women in the cross-sex material suggests that the presence or absence of the opposite sex might be a conditioning factor, but no conclusions can be drawn until a more detailed study is undertaken, distinguishing, for example, different types of tag questions and exploring the possibility of idiosyncratic usage (cf the inconclusive result for the functionally related Swedish particle väl in Einarsson 1978:105).

Softeners or 'pragmatic particles' have a variety of functions, ranging from attitudinal signals to mere 'fillers' (see Crystal & Davy 1975:91ff, Östman 1981, 1982), and they are consequently difficult to generalize about. Their primary role, however, is pragmatic and stylistic rather than semantic, and their informal 'softening' effect - whether they act as hedges or monitoring devices, express the speaker's attitude, or merely serve to keep the conversation going - is indisputable. Without them, the conversation becomes flat and formal. Judging from stereotypes of female speech as 'polite', 'hesitant' and 'accommodating', one might expect softeners to be more frequent in women's speech, but this is contradicted by the results of the cross-sex material, where the men were found to employ softeners 1.5 times as often as the women (28 vs 19 tokens per 1 000 words). The frequent use of certain softeners in the material is well illustrated in the following (male) example:

402 «I mean» I mean she's ‖so ΔLĬTTLE‖ 403 ‖I mean ▷you you ‖KNÓW‖ 404 sort of ‖one can IMΔĂGINE‖ 405 a sort of ‖middle-aged ΔWŎMAN‖ 406 with a ‖coat that 'seemed · you ‖KNÓW sort of‖ · 407 ‖just slightly ex'aggerated her 'ΔFÔRM‖ · 408 you know «I mean» she could sort of ‖slip 'things in ▾ainside ΔPŎCKETS‖ (LLC: S.2.13)

However, the male predominance in the material was restricted to two softeners, sort of and you know, while the use of well, I mean, you see and kind of was fairly evenly distributed between the sexes. The picture is further complicated by the findings of two other students (G. Barkenius and B. Ahman) who made independent studies of the use of softeners in same-sex conversations. These corroborated the male predominance of you know, but revealed a reversed female preference for sort of. Moreover, of the total number of you know in the male material, 86% were uttered by one speaker. This indicates that the use of softeners may be idiosyncratic rather than sex-related, or subject to functional and social differentiation, as suggested by Östman (1981:70ff). On the whole, the use of softeners and other discourse items in spoken English (eg really - see Stenström 1983) is a complex and comparatively unexplored area which deserves more attention in future research.

In face-to-face conversation, paralinguistic signals (gestures, facial expression, tone of voice, etc) are very important. These tend to be sex-differentiated, a fact which has been related to power contrasts like male dominance and authority vs female submission and politeness (eg Thorne & Henley 1975:15ff). In Rosmark's study of hesitation features the female silent pauses were found to be more frequently accompanied by laughs than the male ones. A closer scrutiny revealed that, on the whole, the women laughed five times more than the men (154 vs 31 tokens). This tendency was noticeable in all 'laughing situations' (single or shared laughter, and laughter at the speaker's own or at other speakers' utterances), but was especially pronounced when a woman was the sole laugher at somebody else's remark (109 female vs 9 male tokens). The consistency of this pattern contradicts the popular stereotype that men are wittier than women, but seems to support the claim that laughing is a role-differentiated behaviour of the sexes (cf Thorne & Henley 1975:15-17).

Another frequent observation is that women use more rising tones than men (eg Kramarae 1982:86, Lakoff 1973). Klang's study of nuclear pitch

direction does not confirm this: no significant differences were found, and the slight sex-related tendencies that could be observed seem to contradict rather than support a female preference for rising tones. Another interesting observation was that of 34 male instances of rising tone in her sample (including complex and compound tones with final rising pitch), 22 (65%) were produced by a single speaker. This indicates that nuclear pitch direction, like the use of softeners, may be subject to individual rather than sex-determined variation. Prosodic differences between the sexes no doubt exist, but these seem to involve other intonational properties than nuclear pitch direction alone (see Brend 1975).

These studies of linguistic sex-differentiation in the ETOS material are of course too limited to prove anything about men's and women's language, in English or in general. However, they are suggestive in several ways. They show that popular assumptions and intuitive observations are often badly founded and that they have to be revised or qualified when confronted with genuine data. They also show that, even in areas where some empirical work has been done, our knowledge of male and female speech is still limited and superficial, and that a great deal of further research is needed to clarify the relationship between language and sex.

## 4. CONCLUSION

Although the undergraduate studies presented here have been restricted in various ways, and have differed widely in character, some general conclusions can be drawn. The lexical studies, focusing on different aspects of English vocabulary, have all confirmed the existence of two divergent modes of communication, a spoken 'verbal' style and a written 'nominal' style. The contrast between these has been shown to be systematic and fundamental: it involves categories and structures at all levels of linguistic organization, and it is closely related to the maximally contrasted situations characterizing spontaneous face-to-face conversation and public expository prose. The results of these studies have been 'positive' in the sense that they have confirmed and enriched each other from widely different viewpoints, and agreed well with the findings or other comparable studies. Hence, they seem to have great generality, extending far beyond the specific data on which they have been based.

By contrast, the results of the sex-oriented studies appear to be more 'negative' and disappointing. Although they have sometimes been in agree-

ment, they have more often revealed a confusing picture of non-significant or conflicting tendencies. Yet, in an important sense these results, too, have been 'positive'. They have demonstrated that language and sex is a very complex area where much is assumed but little still is known, and where more research is needed to disentangle the many variables that determine conversational behaviour.

In both these senses - positive as well as negative - undergraduate studies of the kind reported here can be of great value and contribute to a better understanding of various aspects of spoken and written English.

## 5. REFERENCES


Akinnaso, F. Niyi. 1982. "On the differences between spoken and written language". Language and Speech 25. 97-125.

Altenberg, Bengt. Forthcoming. "Causal connection in spoken and written English". Studia Linguistica 38.

Altenberg, Bengt & Gunnel Tottie. This volume. "Will there be texts in this class? Writing term papers within a research project".

Brend, Ruth M. 1975. "Male-female intonation patterns in American English". Language and sex: Difference and dominance, ed. by Barrie Thorne & Nancy Henley, 84-87. Rowley, Mass.: Newbury House.

Chafe, Wallace L. 1982. "Integration and involvement in speaking, writing, and oral literature". Spoken and written language: Exploring orality and literacy, ed. by D. Tannen, 35-53. Norwood, N.J.: Ablex.

Crystal, David & Donald Davy. 1975. Advanced conversational English. London: Longman.

DeVito, Joseph A. 1966. "Psychogrammatical factors in oral and written discourse by skilled communicators". Speech Monographs 33. 73-76.

— 1967. "Levels of abstraction in spoken and written language". Journal of Communication 17. 354-61.

Einarsson, Jan. 1976. "Män, kvinnor och språk". Nordiska studier i filologi och lingvistik. Festskrift tillägnad Gösta Holm, 59-75. Lund: Studentlitteratur.

— 1978. Talad och skriven svenska. Lund: Studentlitteratur.

Elert, Claes-Christian. 1979. "Forskningsfältet 'Språk och kön' vid den Nionde världskongressen i sociologi i Uppsala 1978. Presentation, referat och synpunkter". Könsroller i språk 3. Rapport nr 75 från avdelningen för forskning och utbildning i modern svenska (FUMS), 18-43. Uppsala: Institutionen för nordiska språk.

Forsheden, Oscar. 1983. "Studies on contraction in the London-Lund Corpus of Spoken English". ETOS Report 2. Department of English, Lund University.

Johansson, Stig, Geoffrey Leech & Helen Goodluck. 1978. Manual of information to accompany The Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers. Department of English, Oslo University.

Kramarae, Cheris. 1982. "Gender: How she speaks". Attitudes towards language variation, ed. by Ellen Bouchard Ryan & Howard Giles, 84-98. London: Edward Arnold.

Kroch, Anthony S. & Donald M. Hindle. 1982. A quantitative study of the syntax of speech and writing. Final report to the National Institute of Education on Grant no. G78-0169. University of Pennsylvania, Philadelphia.

Lakoff, Robin. 1973. "Language and woman's place". Language and Society 2. 45-79.

Leech, Geoffrey & Jan Svartvik. 1975. A communicative grammar of English. London: Longman.

Linnarud, Moira. 1979. "Lexikal täthet och betydelsesamband". Svenska i invandrarperspektiv, ed. by Kenneth Hyltenstam, 219-30. Lund: Liber Läromedel.

Ochs, Elinor. 1979. "Planned and unplanned discourse". Syntax and semantics 12: Discourse and syntax, ed. by T. Givón, 51-80. New York: Academic Press.

O'Donnell, Roy. 1974. "Syntactic differences between speech and writing". American Speech 49. 102-110.

Poole, Millicent E. & T.W. Field. 1976. "A comparison of oral and written code elaboration". Language and Speech 19. 305-12.

Stenström, Anna-Brita. 1983. "What does really really do? Strategies in speech and writing". Department of English, Lund University. Mimeo.

Svartvik, Jan. 1980. "Well in conversation". Studies in English linguistics for Randolph Quirk, ed. by Sidney Greenbaum, Geoffrey Leech & Jan Svartvik, 167-77. London: Longman.

Svartvik, Jan, Mats Eeg-Olofsson, Oscar Forsheden, Bengt Oreström & Cecilia Thavenius. 1982. Survey of Spoken English. Report on research 1975-81. Lund Studies in English 63. Lund: CWK Gleerup.

Svartvik, Jan & Randolph Quirk (eds.). 1980. A corpus of English conversation. Lund Studies in English 56. Lund: CWK Gleerup/Liber.

Thavenius, Cecilia. This volume. "Pronominal chains in English conversation".

Thorne, Barrie, & Nancy Henley. 1975. "Difference and dominance: An overview of language, gender, and society". Language and sex: Difference and dominance, ed. By Barrie Thorne & Nancy Henley, 5-42. Rowley, Mass.: Newbury House.

Tottie, Gunnel, Bengt Altenberg & Lars Hermerén. 1983. English in speech and writing/Engelska i tal och skrift. A manual for students. ETOS Report 1. The Departments of English, Lund and Uppsala Universities.

Trudgill, Peter. 1974. Sociolinguistics. Harmondsworth: Penguin Books.

Ure, Jean. 1971. "Lexical density and register differentiation". Applications in linguistics. Selected papers of the Second International Congress of Applied Linguistics, Cambridge 1969, ed. by G.E. Perren & J.L.M. Trim, 443-52. Cambridge: Cambridge University Press.

Vachek, Joseph. 1980. "Remarks on the typology of modern English". Studies in English linguistics for Randolph Quirk, ed. by Sidney Greenbaum, Geoffrey Leech & Jan Svartvik, 271-276. London: Longman.

Wells, Rulon. 1960. "Nominal and verbal style". Style in language, ed. by Thomas A. Sebeok, 213-220. Cambridge, Mass.: The M.I.T. Press.

Zimmerman, Don H. & Candace West. 1975. "Sex roles, interruptions and silences". Language and sex: Difference and dominance, ed. by Barrie Thorne & Nancy Henley, 105-29. Rowley, Mass.: Newbury House.

Östman, Jan-Ola. 1981. You know: A discourse functional approach. (Pragmatics & Beyond II:7.) Amsterdam: John Benjamins.

— 1982. "The symbiotic relationship between pragmatic particles and impromptu speech". Impromptu speech: a symposium, ed. by Nils Erik Enkvist, 147-77. Publications of the Research Institute of the Åbo Akademi Foundation, 78. Åbo: Åbo Akademi.